

Appendix A

Example 1		
Array 1	Array 2	Diff 1_2
a	b	1d0
b	a	< a
aa	bb	2a2,3
bb	aa	> a
aaa	ddd	> bb
bbb	ccc	3a5,6
ccc	bb	> ddd
ddd	aaa	> ccc
xxxx	yyyy	6,9d8
yyyy	xxxx	< bbb
		< ccc
		< ddd
		< xxxx
		10a10
		> xxxx

Example 1 above shows Arrays 1 and 2, representing program elements from two source code files, that have the same lines sorted by line length but with lines of equal length sorted differently. Arrays 1 and 2 were compared using the diff utility referenced in Arnow. Diff 1_2 shows the result of the comparison with 5 non-matching lines in each array, as illustrated with indicators “>” and “<”. The diff utility found these differences because it performed the comparison based on the current order of lines in the arrays, rather than a line-by-line comparison irrespective of the order of lines. Example 1 illustrates that the diff utility fails to identify plagiarism in arrays that have the same lines but a different order of lines, i.e., the problem addressed by the present invention.

Example 2		
Array 1	Array 3	Diff 1_3
a	a	1a2,3
b	a	> a
aa	a	> a
bb	b	2a5,16
aaa	b	> b
bbb	c	> c
ccc	c	> c
ddd	c	> c
xxxx	c	> c
yyyy	c	> c
	c	> c
	c	> c
	d	> d
	d	> d
	e	> e
	f	> f
	aa	4a19,22
	bb	> cc
	cc	> dd
	dd	> ee
	ee	> ff
	ff	8a27,28
	aaa	> eee
	bbb	> fff
	ccc	
	ddd	
	eee	
	fff	
	xxxx	
	yyyy	

Example 2 above shows Arrays 1 and 3, representing program elements from two source code files. Array 3 includes all lines from Array 1, as well as a number of additional lines. Arrays 1 and 3 were compared using the diff utility referenced in Arnov. Diff 1_3 shows the result of the comparison. In particular, the diff utility found a large number of non-matching lines, as indicated by “>” or “<” in Diff 1_3. Example 2 illustrates that the diff utility fails to identify plagiarism in arrays where one array constitutes a complete subset of another array, i.e., the problem addressed by the present invention.